91-03650/

January 1991 DCIEM No. 91-13

An Incentive Learning Approach to Decision Making

T.R.H. Cutmore, DCIEM
R.J. Beninger, Queen's University at Kingston

Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West, P.O. Box 2000
North York, Ontario M3M 3B9

DEPARTMENT OF NATIONAL DEFENCE - CANADA

Table of Contents

Abstract i

1. Introduction 1

2. Incentive Learning 2
   2.1 Definition and Brief Historical Perspective 2
   2.2 Neuronal Model 2
   2.3 Computational Model 4
   2.4 Experimental Basis of Model Elements 7

3. IL and Planning and Decision Making 8
   3.1 Problem Domain 8
   3.2 Deciding 9
   3.3 Planning 10
   3.4 Generalization 10

References 14

Abstract

Incentive learning (IL) theory describes a process whereby an organism learns to orient towards, approach and manipulate certain stimulus objects in the environment. More generally, IL is a process which leads to a decision making capability. The history of development of this theory has its origins in the work of behaviourists such as Hull and Spence in the middle part of this century. More recently, IL theory has been considerably refined by advances in the neurosciences with the study of the biochemical basis of learning. In particular, the neurotransmitter dopamine has been implicated as an important mediator of IL in the brain. The importance of dopamine and IL in human behaviour is indicated by the significant impairments evident in schizophrenia and the effects of neuroleptic drugs. A neuronal model is presented which summarizes these recent developments. A computational model is then outlined which gives a formal description of the neuronal model. Finally, it is suggested that IL theory may be useful in understanding human decision making and furthermore that the computational model may be an important tool for designing and implementing computer programs which have decision making capabilities.

1. Introduction

Planning and decision making are critical activities in a wide range of contexts. In the military, command and control (C^2) depends upon the expeditious use of limited resources in a timely manner requiring effective planning and decision making. General conceptual architectures for the implementation of cognitive processes, for example, ACT* (Anderson, 1983) and SOAR (Laird et al., 1987) and more specific models of decision making in military tasks (e.g., SHOR, Wohl, 1981) have been developed. While existing models put forth some general principles to describe the acquisition of planning and decision making skills, there remains a need to describe and implement a specific mechanism for the acquisition of these skills.

This report describes a possible approach to understanding adaptation to a complex and dynamic environment. The approach takes advantage of recent advances in the neurosciences that provide detailed descriptions of the neuronal mechanisms and molecular events that underlie various forms of learning in a variety of species. This approach may provide a means for actually understanding the mechanisms by which decision making skills are acquired. Successful development and implementation of computer software that executes according to the principles that operate in the brain during planning and decision making activities may offer a powerful new tool for information organization and selection that would provide a significant advantage in C^2.

This report describes some of the first steps that have already been taken by the authors towards the development of a computational model of incentive learning (IL). As described in section 2, IL is a fundamental type of learning that underlies the acquisition, by certain stimuli in the environment, of the ability to elicit approach responses or to attract an animal. This type of learning is essential to effective planning and decision making in humans as evidenced by the severe impairments of these abilities that ensue when the mechanisms underlying IL are dysfunctional, as may be the case in schizophrenia (Cutmore and Beninger, 1990). Section 2.3 presents briefly the computational model being developed for the implementation of IL. Section 3 examines the nature of the decision making process and the application of the IL approach.

2. Incentive Learning

2.1 Definition and Brief Historical Perspective: IL theory describes the process by which environmental stimuli come to be attributed with "importance" or "relevance" as evidenced by the way in which such stimuli engage patterns of behaviour. This type of learning occurs in association with the pursuit of goals that satisfy biological needs. The study of IL has been exclusively restricted to biological organisms, both infra-human and human. However, any system which may have goal oriented states defined upon it may be a candidate for the application of IL mechanisms for guiding its behaviour. IL is not intended to be the singular or autocratic source of behavioural control. However, since it can exert a powerful influence on the behaviour of organisms it may be a useful adjunct for automatic systems. In addition, a formal development of IL theory for an intelligent automatic system may also promote understanding of how a human performs similar cognitive tasks.

IL theory developed within the framework of behaviourism. The classical model of incentive motivation learning was developed by Hull (1952) and Spence (1956). This concept was used to describe the intensity of ongoing behaviour, for example, an organism pursuing a goal with high intensity, as indexed by the speed and vigor of responses. The organism was said to be motivated and the goal which was pursued to have incentive value. The concept of incentive motivation was given a more cognitive interpretation by Bolles (1972) and Bindra (1974). While traditional behaviourism strictly confined itself to directly observed behaviour, these authors suggested the utility in considering internal representations as mediators of behaviour. More recently, Beninger (e.g., 1983, 1990), investigating the neuronal and neurochemical basis of IL, has argued that the neurotransmitter dopamine plays an important role. Knowledge of the neuronal bases of IL permits a more concrete understanding of the nature of the representations referred to by Bindra and Bolles.

2.2 Neuronal Model: Each environmental stimulus encountered by an organism has associated with it a distinct pattern of neuronal activation, the central representation of the stimulus. The organism comes equipped with the capacity to direct a wide range of responses towards environmental stimuli; each response produces a distinct pattern of efference copy and proprioceptive feedback, the central representation of the response. It is well known that stimuli and responses lead to the activation of unique subsets of cortical neurons and that the heaviest

projection area of the cortex is the striatum, including caudate, putamen and accumbens nuclei (cf., Beninger, 1983). The output of the striatum includes pre-motor nuclei in the brainstem that influence motor neurons and the substantia nigra (cf., Smith and Bolam, 1990). Finally, dopaminergic neurons in midbrain nuclei (including the substantia nigra) project diffusely within the striatum (Anden et al., 1966). These neurons have been shown to be activated by the presentation to an animal of a goal object (Blackburn et al., 1986; Keller et al., 1983; Phillips et al., 1989), often a biologically significant stimulus that satisfies some need; such goal objects are defined as primary incentive motivational stimuli or, more conveniently, rewards.

The consequences of finding a reward and activating the dopaminergic neurons is an increase in the ability of environmental stimuli that immediately preceded that event to elicit an approach response in the future, i.e., incentive learning. According to the IL model, the neurons constituting the central representation of the stimulus will have undergone a dopamine-mediated change in their ability to activate striatal output neurons (and therefore response systems) in the future; similarly, the central representation of the response that was ongoing at the time of the reward will undergo incentive learning, increasing the likelihood of that particular response recurring in the future when the conditioned incentive stimulus is encountered. An important feature of the IL model is that a diffuse dopaminergic signal within the striatum in association with reward will lead to a modification of only the cortical-striatal projections that were most recently active. This may be achieved by establishing a transient state of readiness in those projections in conjunction with their activation. If reward is encountered and a dopaminergic signal occurs during the period of readiness, IL will occur at those synapses.

It is important to note that IL is not the only type of learning. Researchers studying IL can block this type of learning with the use of pharmacological agents that decrease dopaminergic neurotransmission (see Beninger, 1983, 1990). However, in these experiments it was found that animals failing to show IL retained the capacity to learn associations among stimuli. This latter type of learning was termed stimulus-stimulus associative learning (SSL) and appeared to be mediated by a different mechanism than IL.

Most rewarding stimuli achieve this status because they satisfy some biological need in an animal. Once the need is satisfied, the reward capacity of a particular stimulus, e.g., food, may be

temporarily reduced. Thus, animals may be said to have motivational states associated with specific needs and these states will have a variable intensity depending on the current level of satisfaction of that need. In the present context, the motivational state would define those objects in the environment that are goals. For this reason motivational states are incorporated into the computational model which follows.

2.3 Computational Model: While the above description provides the basis for IL, there remains a need for a computational approach to describe how to build a system (either in simulation or in actual hardware) that learns via IL. The nature of environmental stimuli, central representations and responses needs a formal description such that a computer program can be developed to realize this objective. Therefore, the first task is to provide a computational model of IL and secondly to create programs to implement it. Cutmore and Beninger (in preparation) describe elements of the model which are outlined here.

IL is supported by SSL which structures the associations between three broad classes of central representations: environmental stimuli (E_s), motivational states (I_s) and response efference (R_s). SSL is responsible for constructing a cognitive map of the environment in E_s. It may also result in associations between central representations which indicate the motivational states of the organism in I_s. In addition, SSL may result in response sequence development so that the sequence as a whole may be treated as though it were a unitary, uninterrupted response in R_s. Finally, SSL may also result in associations among elements of different classes. Thus, an environmentally based sensory representation, E_i, may evoke a representation in Is (for example, the feeling of hunger) or evoke an element of R_s (for example, an intention to move towards the E_i). In this way SSL may yield a network of associations which provides the context or knowledge about a situation; IL selects from these associations and provides a focus for action.

An environment exists as an ordered collection of stimulus elements which may be designated by their location as an n-tuple in n-dimensional space:

$$E^n = \{ X_n \mid X_n = (x_1, x_2, .. x_i, .. x_n);\ 1 \leq n \leq Z;\ Z \in N \} \qquad (1)$$

An element of E^n is given by E_{ijk}. Input to the system is given by a function which yields a subset of elements from E^n, to create pattern E_{ij}, a particular instance of a sensory input. This pattern

corresponds to the central representation of a stimulus referred to above. E_i denotes a set of stimulus patterns which are similar, for example the set of views of some object in the environment. The set of all E_i defines E_s.

The central representation of a motivational state consist of elements Iijk. An instance of a particular motivational state is Iij and the set of similar motivational states is grouped functionally into I_i. The set of all I_i defines I_s.

Central representations of response classes (R_i) may be interpreted as tendencies to perform certain types of actions, such as, "move towards object X". They may be derived from previous experience via efference copies of past behaviour and proprioceptive feedback. A response representation may be realized in behaviour by stimulating a motor command (C_i). A particular instance of the class, R_i, is denoted, R_{ij}, for example, "Move toward object X by walking". The neural elements of an instance, R_{ij}, are denoted by R_{ijk}. The set of all R_i defines R_s.

In each of the above cases, classes of domain elements are referred to by a single indexed variable, such as E_i to denote the set of all sensory patterns that represent an object. While the computations will generally take place at the level of specific patterns (double-indexed variables) and pattern elements (triple-indexed variables), it will simplify further discussion to refer mainly to classes. The level of discussion is therefore in terms of objects, goals and actions. In equations where a function range variable occurs as a parameter of the function domain, the time variable 'j' is used to distinguish them.

S_i is a complex consisting of a member from each of E_s, I_s and R_s. There are two important properties of an S_i for the model. One of these properties has been acquired and structured by SSL. SSL results in one representation being capable of eliciting another representation, either within the same domain, or within a different domain. Each component of S_i may therefore elicit representations from each of the domains. In particular, elements in the response representation domain may be stimulated. This would give rise to a set of tendencies to act at time j+1 according to†:

$$G: A(S_i)_j \text{ X } w(S_i,R_i)_j \text{ X } A(R_i)_j \rightarrow A(R_i)_{j+1} \qquad (2)$$

---

†. The Cartesian product is indicated with 'X' in function definitions.

where $A(S_i)$ is the current activation of Si, $w(S_i,R_i)$ is the associative strength between S_i and R_i, and $A(R_i)_j$ is the current activation of R_i, yielding a new activation for R_i, $A(R_i)_{j+1}$. In general, the activation of a neural pattern decays in time if not re-stimulated and thus there are temporal effects involved in computing G.

The second property is the incentive value of S_i, termed S_i*, and is acquired by incentive learning (IL). S_i sends widely divergent efferents onto command neurons (C) with associative strength $w(S_i,C_i)$. Prior to incentive learning, w has moderate strength and corresponds to the "novelty" of a stimulus. It is computed by: $G(A(S_i), w(S_i,C_i), A(C_i))$. S_i becomes more or less effective in stimulating a subset of C as incentive learning proceeds. Before discussing how this occurs it should be noted that a major distinction between SSL and IL is in the mechanisms which modify w. The processes involved in SSL may occur according to Klopf(1988), involving contiguity factors and will not be covered here. An important property of both forms of learning is that w must be permitted to both increase and decrease but the mechanism proposed for IL differs markedly from those proposed for SSL.

For IL, the connection strength is modified as:

$$J: R(S_i)_j \times D_j \times A(S_i)_j \times w(S_i,C_i)_j \rightarrow \Delta w(S_i,C_i)_{j+1} \qquad (3)$$

The magnitude of the change is proportional to the state of "readiness", $R(S_i)$, induced by the stimulus. This quantity follows a time function with a maximum at some time after S_i is presented. D is a reward which corresponds to the magnitude of the dopamine signal to the striatum noted in section 2.2. The w term in eqn. (3) is used to compute a reduction factor due to the use of a synapse in proportion to $A(S_i)$. Thus, unrewarded presentation of S_i results in negative $\Delta w(S_i,C_i)$.

The reward signal is generated by a subset of central representations which have incentive value, and is given by:

$$* : A(S_i) \times w(S_i, C_i) \times A(C_i) \rightarrow D. \qquad (4)$$

The incentive value of a stimulus (S_i*) is given its definition by this function. This corresponds to the collateral feedback from striatal output neurons to the reward source mentioned in section 2.2.

Finally, the probability of some action B_i (which is equivalent to a set of behaviours with the same effect) occurring is jointly determined by applying eqn. (2) to compute $A(R_i)$ by SSL and applying eqn. (2) to compute the generalized activation of the command subset $A(C_i)$. K is a probability function which takes these parameters into the set of possible actions B:

$$K: A(R_i) \text{ X } A(C_i) \rightarrow p(B) \qquad (5)$$

yielding a probability density over B which ascribes a likelihood to each action: $p(B_i)$.

For behaviour to be flexible and effective, an action must be sensitive to the environmental context and the motivational state of the system. Since E_i and I_i are subcomponents of S_i, $p(B_i)$ will be sensitive to environmental context and motivational state. It is in this way, that stimuli of E_s and motivations of I_s may operate as discriminative stimulus components of S_i. In addition, it was noted above that Is may also be structured by SSL (e.g., by temporal contiguity) such that the satisfaction of one goal may be associated with the activation of another one. Similarly, E_s is structured to form a cognitive map of the environment and thus the current stimulus representation may anticipate the input of other stimuli. These associative structures in E_s and I_s may provide important control of the state of S_i and thus behaviour.

It is also important that actions occur in the proper sequence. Actions may be appropriately sequenced by two processes: (1) Sequences of behaviours will occur as a consequence of the cycle of computation. S_i also has an R_i component from the previous response representation which resulted in action $B_{i(j)}$. This will partially control the selection and execution of $B_{i(j+1)}$. A subsequent reward signal (or the lack thereof) will alter the probability of $B_{i(j+1)}$ occurring under similar stimulus circumstances in the future and thus the likelihood of the action sequence: $<B_{i(j)},B_{i(j+1)}>$. (2) As incentive learning proceeds and behaviours are brought into temporal sequence, SSL associations may operate to form sequence chains within R_s, which may then be treated as individual R_i. In this way, by practice after IL has sequenced behaviours by a computationally intensive cycle, behaviour sequences may be executed more rapidly, yet with less interruptive control, to yield skilled action sequences.

2.4 Experimental Basis of Model Elements: It is important to point out that the elementary quantities of the computational model [viz., A(S), A(R), A(C), R(S) and w] can each be

substantiated in terms of behavioural and neurobiological experimentation. The A(S) term stands for the activation of a stimulus. The effects of manipulating a stimulus pattern (e.g., on a measure of similarity) can result in a gradient of responding in a discrimination paradigm. In addition, a change in the stimulus intensity for a constant pattern can result in correlated changes in response strength (Mackintosh, 1974). The A(R) term stands for the activation of an image or efference copy of a previously executed response. Among the experimental methods, a delay of reward paradigm has been used. For example, it has been shown that responses which have longer delays between execution and reward presentation are more difficult to train (Mackintosh, 1974). In the model, this corresponds to a decay in the efference representation, A(R), over time. The A(C) term stands for the degree of activation of command output which activates motor systems. This quantity can be manipulated pharmacologically to produce biases in motor output, such as turning behaviour (Pycock, 1980). The R(S) term stands for the transient state of "readiness" of synapses activated by a stimulus pattern. Early brain stimulation reward studies (e.g., Stein, 1958) and more recently experiments on conditioned activity (Mazurski and Beninger, 1991) have illustrated this synaptic property indirectly through manipulations of DA. Finally, experimental support for w, which stands for the connection strength between two neural patterns, has been illustrated by many behavioural studies (for example, through the manipulation of the number of pairings of stimuli in a variety of paradigms such as classical conditioning, sensory preconditioning and operant conditioning). Pharmacological studies have examined the synaptic basis of neurotransmitter mediated changes in synaptic plasticity (e.g., in studies of Aplysia, Kandel and Schwartz, 1985; and with DA in higher organisms, Ehlert et al., 1981) which provide evidence of a physical substrate for w and changes in this quantity with learning.

3. IL and Planning and Decision Making

3.1 Problem Domain: Typically, E^n offers a variety of stimuli which may be selected and acted upon. The organism also typically has multiple (and possibly competing) motivations which may be active simultaneously leading to a set of goals which vary in their mutual compatibility. In other words, the organism will have various options for action with some optimal action sequence which

depends upon the environmental context and a complex motivational state. A decision mechanism is required to determine which action should be performed at each point in time. However, in a complex environment responding in accordance to current circumstances only is insufficient. The principle reason is that prevailing circumstances under-determine which action will be effective (rewarding) in the long run. Therefore, a means of planning a sequence of future actions is required. In other words, it is often the case that the acquisition of a reward will require many steps and making some choices limit or prohibit the range of subsequent effective choices.

3.2 Deciding: IL occurs as a consequence of experience in dealing with the environment. In conjunction with SSL it provides a system with the means of making decisions based on current circumstances. In addition, a system equipped with these learning schemes is capable of some limited forecasting of the consequences of an action. This occurs because of the operation of SSL using temporal contiguity as the basis of association formation. Consider the following example: A travelling salesperson, M, drives a car through an unfamiliar town enroute to some other town. After a sequence of turns, made with some trial and error (as there are no road signs) the way is made, and a reward signal occurs ("whew!, out of there at last"). Upon returning through the area the following week M follows a somewhat different route and a sequence of turns is made which results in a quicker thoroughfare, a rewarding result. This occurs several times, with some better fares and some worse, until an optimal route is found. We may assume that each potential turning point has some distinguishing feature (E_i). IL can explain the route learning as follows: Initially, each E_i has equal significance and thus a left-turn or right-turn has similar probabilities (admittedly they will not, as moves are made and circling is avoided). After a sequence of turns, M exits the town. When the reward signal occurs, those Si complexes which were recently active acquire incentive value. This means that when they are encountered in the future as M navigates the town again, they will activate general response systems (commands to act, i.e., to make a turn). Which behaviour is executed will depend upon which response representation has been associated with that S_i in the past. Thus, a dual contingency leads M to make a decision: (1) general activation of response systems to act and (2) the retrieval and biasing by SSL of the particular response to make. Deciding what to do turns out to depend on both knowing one has to act (being aroused to action) and knowing which response to make. If either of these is absent or deficient then the decision is

impaired.

3.3 Planning: There are two ways in which "planning" may occur in this simple situation. When a sequence of moves is made they cause a temporal sequence of E_i to occur. SSL can associate these by temporal contiguity such that upon input of one E_i the others may be elicited by an associative "chain reaction", a kind of mental navigation of the town or more generally, a mental navigation of the space, E^n. In this way M is provided with one means of forecasting the consequences of making a decision at some point in the route. If the terminal stimulus in the chain is an "exit" E_i, then the sequence of actions to realize this mental route could be performed. Perhaps several associative chains could be examined and then the best one selected. If the associative links between E_i alone are too weak then a second mechanism may operate to plan a sequence of moves. Response representations (R_i) can be associated with perceptual inputs (E_i) by their co-occurrence during the trial and error learning to form a simple two component sequence: $<R_i,E_i>$. In other words, M may anticipate an E_i by it being evoked by an active (imagined) R_i. For example, one may anticipate what one would see if a left-turn were executed at some point in E^n. Furthermore, E_i may also elicit response representations to suggest the next action, yielding the two component sequence $<E_i, R_i>$. This may be regarded as a form of recall of the action one performed the last time one was in that part of E^n. By performing several $<R_i, E_i> <E_i, R_i>$ by this somewhat laborious process, an "exit" E_i may be encountered (i.e., a stimulus with incentive value). A solution sequence of moves has been found and may be considered among others as a sequence to act upon. This final point implies a dual process model of planning and is consistent with experimental evidence. Anderson (1982,1987) has done empirical work on the acquisition of cognitive skill in human subjects and has provided a dual process symbolic information processing model. The neural computation model presented here differs in that it provides a more detailed description of the atomic processes which may underlie the acquisition process.

To summarize, the above example serves to illustrate the following points: (1) IL strengthens the ability of certain stimuli to activate response systems to arouse a system to act. (2) SSL builds a set of associations between central representations which can bias which action is performed. If this leads to a solution (reward signal) then (1) applies again and so on, iteratively. (3) decision

making may be construed as the joint action of (1) and (2); i.e., knowing when to act and how given a current situation. (4) Planning of a future sequence of actions may arise as a result of either a simple associative "chain-reaction" of transformations in E^n based on past experience or by a more laborious, response, change of input, response etc. mental sequence which has been structured by past IL and SSL jointly.

3.4 Generalization: Since E^n has been defined to be an n-dimensional space, it could be applied more broadly than as a source of environmental stimuli. More generally, E^n may be interpreted as a problem space (cf., VanLehn, 1989). R_i and their associated commands and actions could then be considered operators which change the location of a system navigating this problem space. SSL serves to map out the topography of this space while IL attributes importance to specific elements within it as being partial solutions, just as reaching some marker in navigating the town can provide an indication of a partial solution to the problem of getting to the "exit" point in a physical interpretation of E^n. The model proposed here, therefore, has a potential for providing insight into human problem solving. In particular, it may provide an effective means of explaining and therefore simulating some aspects of human planning and decision making. Finally, practical applications of the model in the context of military C^2 would be as a a tool for understanding the acquisition of decision making capabilities in human subjects and the design and development of computer-based decision aids.

References

Anden, N.E., Fuxe, K., Hamberger, B., & Hokfelt, T. (1966) A quantitative study of the nigro-neostriatal dopamine neuron system in the rat. *Acta Physiol Scand, 67*, 306-312.

Anderson, J.R. (1982) Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, J.R. (1983) *The Architecture of Cognition*. Harvard Univ. Press: Cambridge.

Anderson, J.R. (1987) Skill acquisition: compilation of weak-method problem solutions. *Psychological Review, 94*, 192-210.

Beninger, R.J. (1983) The role of dopamine in locomotor activity and learning. *Brain Res Rev,* 6 173-196.

Beninger, R.J. (1990) Receptor subtype-specific dopamine agonists and antagonists and conditioned behaviour. In P. Willner & J. Scheel-Kruger (eds) *The Mesolimbic Dopamine System: From Motivation to Action*. John Wiley and Sons Ltd: Chichester.

Bindra, D. (1974) A motivational view of learning, performance and behavior modification. *Psych Rev, 81*, 199-213.

Blackburn, J.A., Phillips, A.G., Jakubovic, A.., & Fibiger, H.C. (1986) Increased dopamine metabolism in the nucleus accumbens and striatum following consumption of a nutritive meal but not a palatable non-nutritive saccharine solution. *Pharmacol Biochem Behav 25,* 1095-1100.

Bolles, R.C. (1972) Reinforcement, expectancy and learning. *Psych Rev, 5* 394-409.

Cutmore, T.R.H., & Beninger, R.J. (in preparation) A computational model of incentive learning.

Cutmore, T.R.H., & Beninger, R.J. (1990) Do neuroleptics impair learning in schizophrenic patients? *Schizophrenia Res*, 3 173-186.

Ehlert, F.J., Roeske, W.R. & Yamamura, H.I. (1981) Striatal muscarinic receptors: Regulation by dopaminergic agonists. *Life Sci, 28* 2441-2448.

Hull, C.L. (1952) *A Behavior system*. Yale Univ. Press: New Haven.

Kandel, E.R. & Schwartz(1985), J.H. *Principles of Neural Science* 2nd Ed. Elsevier: New York.

Keller, R.W., Stricker, E.M., & Zigmond, M.J. (1983) Environmental stimuli but not homeostatic challenges produce apparent increases in dopaminergic activity in the striatum: an analysis by in vivo voltammetry. *Brain Res, 279,* 159-170.

Klopf, A.H. (1988) A neuronal model of classical conditioning. *Psychobiology, 16,* 85-125.

Laird, J.E, Newell, A. & Rosenbloom, P.S. (1987) Soar: An architecture for general intelligence. *Artific Intell, 33*, 1-64.

Mackintosh, J.J. (1974) *The Psychology of Animal Learning*, Academic Press: New York.

Mazurski, E.J. & Beninger, R.J. (1991) Effects of selective drugs for dopaminergic D-1 and D-2 receptors on conditioned locomotion in rats. *Psychopharmacology* (Berlin), in press.

Phillips, A.G., Blaha, C.D., & Fibiger, H.C. (1989) Neurochemical correlates of brain stimulation reward measured by ex vivo and in vivo analyses. *Neurosci Biobehav Rev, 13*, 99-104.

Pycock, C.J. (1980) Turning behavior in animals. *Neuroscience 5*, 461-514.

Smith, A.D. & Bolam, J.P. (1990) The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurons. *Trends in Neurosci, 13*, 259-265.

Spence, K.W. (1956) *Behavior Theory and Conditioning*. Yale Univ. Press: New Haven.

Stein, L. (1958) Secondary reinforcement established with subcortical stimulation. *Science, 127*, 466-467.

VanLehn, K (1989) Problem solving and cognitive skill acquisition. In: M. I. Posner, Ed. *Foundations of Cognitive Science*, MIT Press.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 12.)

Defence and Civil Institute of Environmental Medicine
1133 Sheppard Ave W. North York, Ont. Canada

2. DOCUMENT SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable)

Unclassified

3. DOCUMENT TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C,R or U) in parentheses after the title.)

An Incentive Learning Approach to Decision Making

4. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Interim research report

5. AUTHOR(S) (Last name, first name, middle initial. If military, show rank, e.g. Doe, Maj. John E.)

Cutmore, Tim R.H. and Beninger, Richard J.

6. DOCUMENT DATE (month and year of publication of document)	7a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)	7b. NO. OF REFS (total cited in document)
	14	25

8a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)	8b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)
Proj #51CJ 13	

9a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)	9b. OTHER DOCUMENT NO.(S) (Any other numbers which may be assigned this document either by the originator or by the sponsor)
DCIEM No. 91-13	

10. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

- (X) Unlimited distribution
- ( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
- ( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
- ( ) Distribution limited to government departments and agencies; further distribution only as approved
- ( ) Distribution limited to defence departments; further distribution only as approved
- ( ) Other

11. ANNOUNCEMENT AVAILABILITY (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (10). However, where further distribution (beyond the audience specified in 10) is possible, a wider announcement audience may be selected.)

Unlimited

12. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

Same as 1

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both offical languages unless the text is bilingual).

Abstract

Incentive learning (IL) theory describes a process whereby an organism learns to orient towards, approach and manipulate certain stimulus objects in the environment. More generally, IL is a process which leads to a decision making capability. The history of development of this theory has its origins in the work of behaviourists such as Hull and Spence in the middle part of this century. More recently, IL theory has been considerably refined by advances in the neurosciences with the study of the biochemical basis of learning. In particular, the neurotransmitter, dopamine has been implicated as an important mediator of IL in the brain. The importance of dopamine and IL in human behaviour is indicated by the significant impairments evident in schizophrenia and the effects of neuroleptic drugs. A neuronal model is presented which summarizes these recent developments. A computational model is then outlined which gives a formal description of the neuronal model. Finally, it is suggested that IL theory may be useful in understanding human decision making and furthermore that the computational model may be an important tool for designing and implementing computer programs which have cedision making capabilities.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible, keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

learning motivation decision making model